



Designing and Developing an Integrated Text Analysis Platform for all levels of Technical Expertise

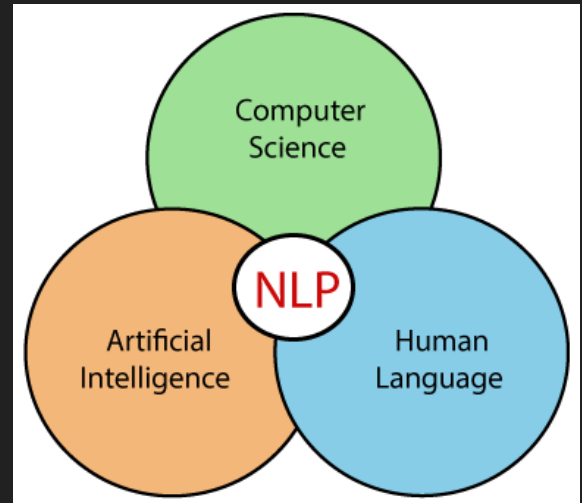
Plan

- What is NLP?
- The Measuring the Earth project
- Design Process & Choices
- What we implemented
- Reflections

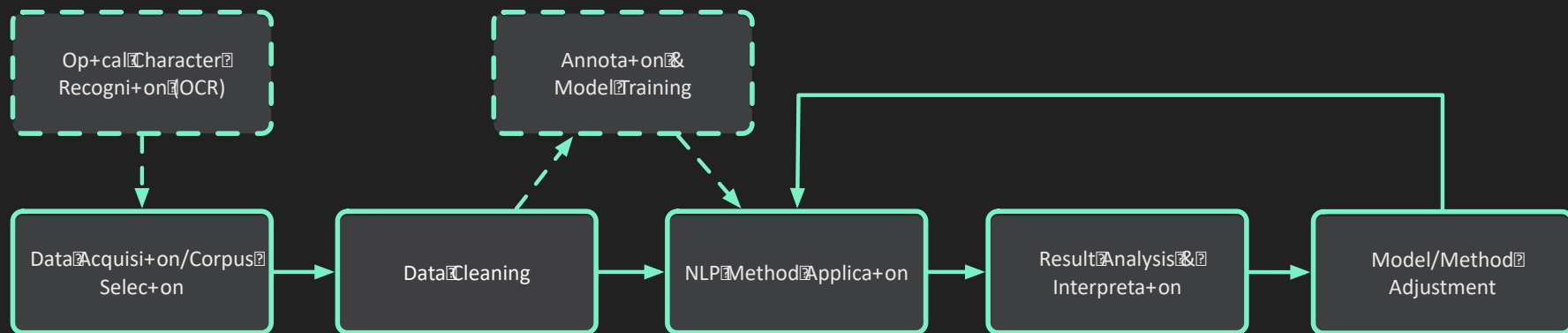
What is Natural Language Processing (NLP)

natural language processing *n.* Computing a form of computational linguistics in which natural-language texts are processed by computer (for automatic machine translation, literary text analysis, etc.); abbreviated *NLP*.

"natural language processing, n." OED Online, Oxford University Press, December 2021, www.oed.com/view/Entry/255272. Accessed 2 February 2022.



The NLP Lifecycle



Building Measuring the Earth



Measuring the Earth

Search

Browse

Batches

Entities

Notebooks

Search

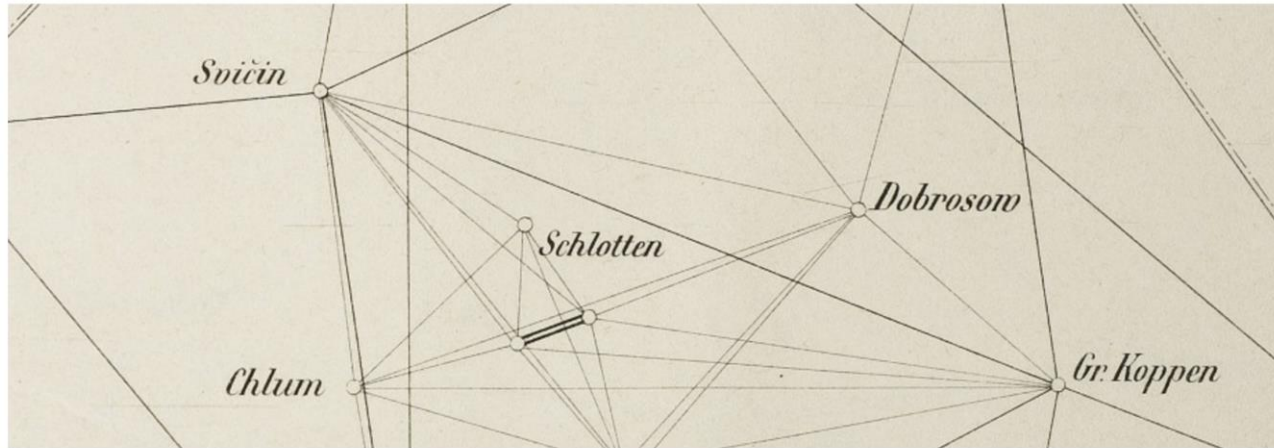
Search

root

Sign out

Measuring the Earth

A Digital Repository for the History of Modern Geodesy

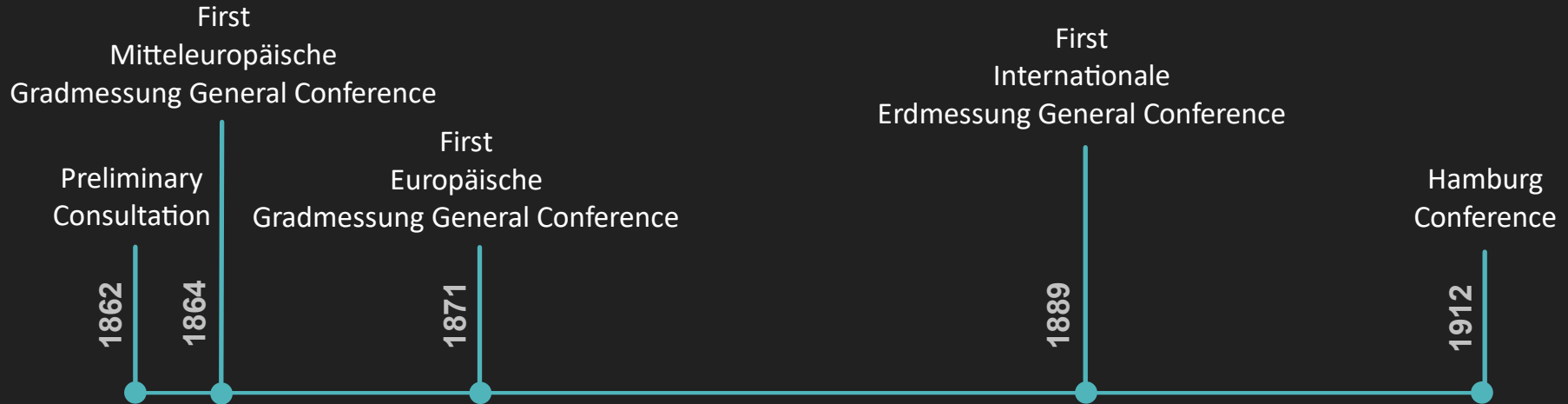


Here, you can explore the reports and minutes produced by the precursors organizations of the current [International Association of Geodesy](#) between 1862 and 1912. The materials have been made available for digital inquiry through a digitization process made possible by a cooperation between the libraries of the [Max Planck Institute for the History of Science](#) and the [GFZ German Research](#)

Measuring the Earth: A DH Project

- Goals:
 - To create an open repository of primary sources for the history of late modern geodetical research
 - To perform a historical analysis of
 - international cooperation
 - scientific diplomacy
 - interdisciplinary debate in the field
- Challenges:
 - OCR
 - Multilingual corpus with older versions of language

The Corpus



Adam, Jozsef & Drewes, Hermann & Rozsa, Szabolcs. (2012). The International Association of Geodesy (IAG) Historical Overview. *Journal of Geodesy*. 86. 793-974.

The Corpus

- High-resolution scans of reports and meeting minutes
- Multilingual (mostly German, French, Italian and English)
- About 15 000 pages

202

VERHANDLUNGEN

DER VOM 20. BIS 29. SEPTEMBER 1875 IN PARIS VEREINIGTEN

15

PERMANENTEN COMMISSION

DER

EUROPAEISCHEN GRADMESSUNG

REDIGIRT VON DEN SCHRIFTFÜHRERN

C. BRUHNS. A. HIRSCH.

ZUGLEICH ALS GENERALBERICHT FÜR DAS JAHR 1875

HERAUSGEGEBEN VOM

CENTRAL-BUREAU DER EUROPAEISCHEN GRADMESSUNG.



COMPTES-RENDUS

DES SÉANCES DE LA COMMISSION PERMANENTE DE

L'ASSOCIATION GÉODÉSIQUE INTERNATIONALE

POUR LA

MESURE DES DEGRÉS EN EUROPE

RÉUNIÉ A PARIS DU 20 AU 29 SEPTEMBRE 1875.

I. 646 b.

RÉDIGÉS PAR LES SECRÉTAIRES

C. BRUHNS. A. HIRSCH.

PUBLIÉS POUR SERVIR DE

RAPPORT GÉNÉRAL POUR L'ANNÉE 1875

PAR LE

BUREAU CENTRAL DE L'ASSOCIATION GÉODÉSIQUE INTERNATIONALE.



BERLIN, 1875.

16

VERLAG VON GEORG REIMER.

Inhalt. — Table des Matières.

Procès-verbaux des séances de la Commission permanente tenues à Paris
du 20 au 29 Septembre 1875.

	pag.		pag.
Première Séance.		nant la liste des publications géodésiques	
Liste des membres de la Commission permanente, des délégués, et des invités . . .	3	et le tableau des déterminations astronomiques	16
Discours d'ouverture de M ^{rs} . Jourdan, Ibañez, Baeyer et Faye	4	Discussion sur ce rapport	18
Programme pour la réunion de la Commission permanente à Paris, en 1875	5	Décisions prises à ce sujet	19
Rapport de la Commission permanente . . .	6	Cinquième Séance.	
Rapport du Bureau Central et de l'Institut géodésique de Berlin	7	Communication d'une lettre de Mr. Fearnley sur les travaux géodésiques en Norvège . . .	20
Seconde Séance.		Rapport de Mr. d'Oppolzer, rapporteur de la Commission spéciale du pendule	20
Rapports des M ^{rs} . d'Oppolzer, de Bauernfeind, Adan, Ibañez, Perrier et Villarceau	10	Résolutions adoptées sur la question du pendule	22
Mr. Breton de Champ, communication sur la différence de niveau entre l'Océan et la Méditerranée, résultant des Nivellements en France, et discussion sur cette question	11	Rapport de Mr. le commandant Perrier, rapporteur de la Commission spéciale pour l'appareil des bases	23
Troisième Séance.		Propositions de la Commission spéciale pour l'appareil des bases, adoptées par la Commission permanente	25
Rapport de Mr. de Vecchi. Lettre de Mr. Stamkart	14	Discussion sur les questions du programme concernant les maréographes	25
Détails donnés par Mr. Baeyer sur l'étude des déviations de la verticale dans le Harz et les environs	15	Propositions de Mr. Hirsch adoptées	26
Rapports de M ^{rs} . Baeyer, Peters, Barozzi, de Forsch, Bruhns et Hirsch	15	Communication de Mr. Hirsch sur l'établissement international des poids et mesures . .	27
Quatrième Séance.		Discussion sur l'horizon fondamental pour les altitudes	29
Rapport de Mr. Mahmoud Bey sur les travaux en Egypte	16	Résolution à cet égard proposée par Mr. d'Oppolzer	30
Rapport de Mr. Hirsch sur l'exécution des décisions prises l'année dernière, concer-		Discussion sur la question des attractions locales et des déviations de la verticale . . .	30
		Discours de Mr. le Ministre de l'Instruction qui assiste à la séance	31

Erste Sitzung.

Paris, im Hôtel des Ministeriums der auswärtigen Angelegenheiten,
den 20. September 1875.

Präsident: Herr General *Ibañez*.

Schriftführer: Herr *Bruhns* und Herr *Hirsch*.

Anwesend von der permanenten Commission: die Herren *Ibañez*, *Baeyer*,
von Bauernfeind, *Bruhns*, *Faye*, *von Forsch*, *Hirsch*, *von Oppolzer*, *de Vecchi*; von den
Commissaren: die Herren *Adan*, *Barozzi*, *Ferrero*, *Perrier*, *Peters*, *Ricci*, *Saget*,
Villarceau.

Eingeladene:

Herr *Jourdain*, Mitglied des Instituts, General-Secretair des Ministeriums des
öffentlichen Unterrichts u. s. w.;

„ *Frémy*, Präsident der Akademie der Wissenschaften in Paris;

„ *Chasles*, Mitglied des Instituts;

„ *Tresca*, Mitglied des Instituts;

„ Commandant *Mouchez*;

„ *Bréquet*, Mitglied des Instituts;

„ *Tisserand*, Director des Observatoriums in Toulouse;

„ General *Billot*, Abgeordneter der National-Versammlung;

„ *Bugnot*, Escadron-Chef im Generalstab;

„ Hauptmann *Bassot*, Correspondent des Bureau des longitudes;

„ *Bouquet de la Grye*, Ingenieur-Hydrograph der Marine;

„ *Alphand*, General-Inspector des Brücken- und Wegebauwes, Director der
öffentlichen Arbeiten in Paris;

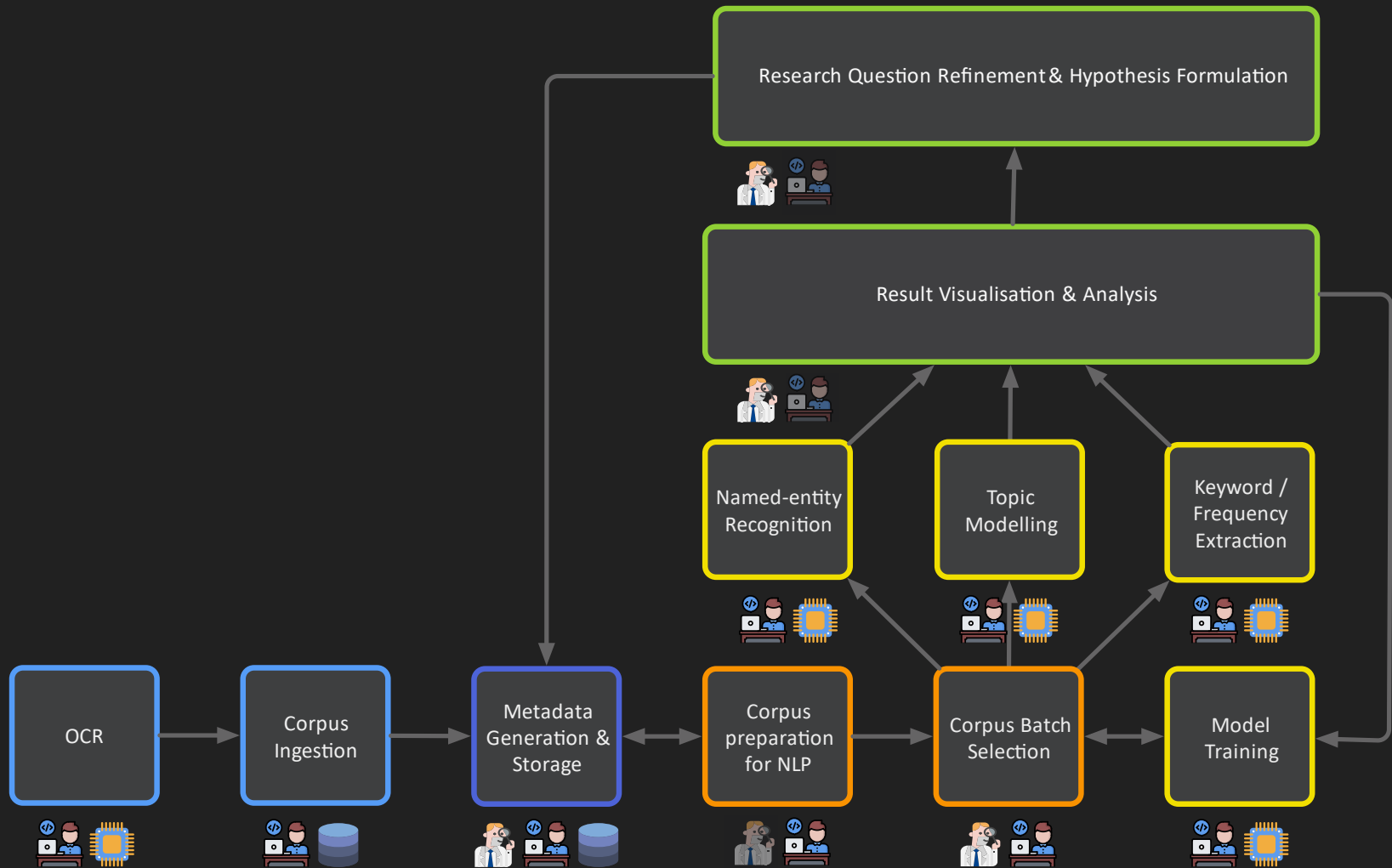
Der Herr *Präsident* erklärt die Sitzung für eröffnet.

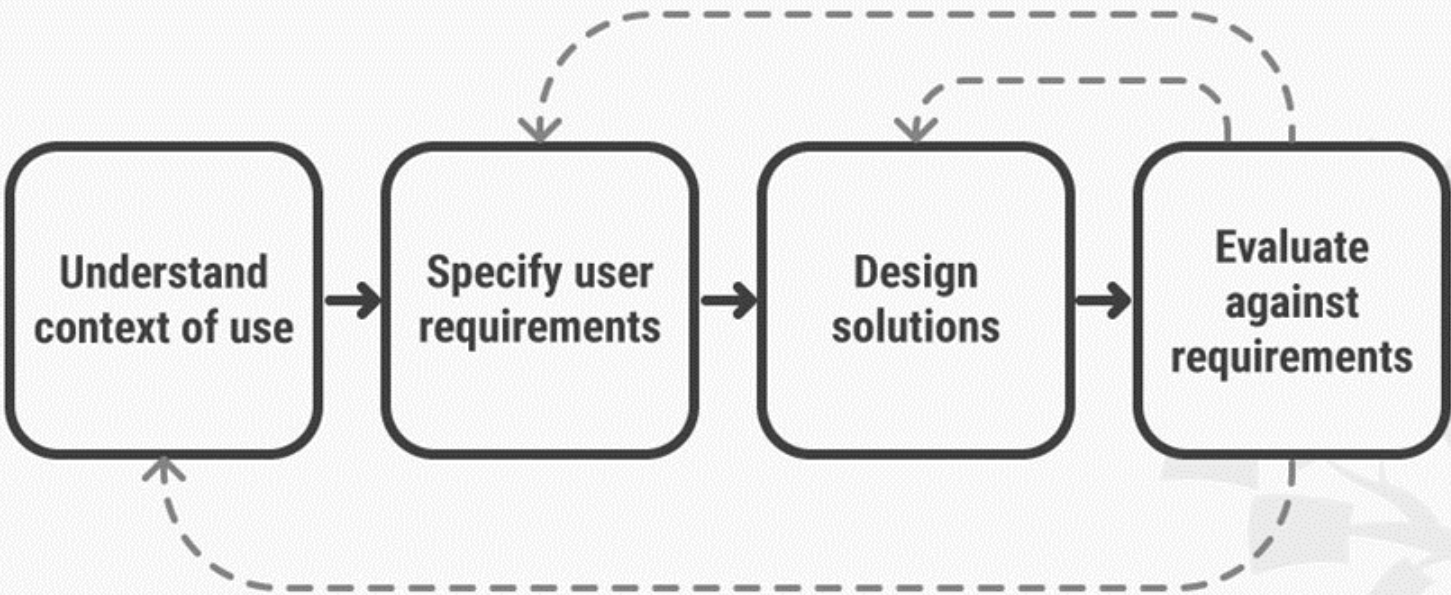
Herr *Jourdain*, Mitglied des Instituts, General-Secretair des Ministeriums des öffentlichen Unterrichts u. s. w., nimmt das Wort im Namen des Herrn Ministers, welcher durch einen bedauerlichen Unfall verhindert ist, der Sitzung, wie er beabsichtigt hatte, beizuwohnen, um der geodätischen Commission zu danken, dass sie der Einladung der französischen Regierung nachgekommen ist, und um derselben zu bekunden, mit welchem Interesse man in Frankreich das von der Commission für die Europäische Gradmessung unternommene wissenschaftliche Werk begrüsst. Frankreich, sagt der Herr General-Secretair, bleibt seinen alten Traditionen getreu, indem es an Forschungen Theil nimmt, welche es seiner Zeit hervorgerufen hat, und indem es sich an einem Unternehmen theiligt, welches durch die Pflege der Wissenschaft und durch das Gefühl europäischer Zusammengehörigkeit erweckt worden ist.

Der Herr *Präsident* antwortet, indem er der französischen Regierung für die sympathischen Gesinnungen dankt, welche sie soeben der geodätischen Commission ausgesprochen hat; die Commission schätzt sich glücklich, in der Hauptstadt des Landes zu tagen, welches man als die Wiege der modernen Geodäsie bezeichnen kann. — Bei der Unterstützung, welche die Regierungen den Bestrebungen der Commission ohne Unterlass gewähren und bei der vortrefflichen Hilfe der Geodäten und Astronomen der verschiedenen Länder kann man hoffen, dass die Geodäsie noch vor dem Ende unsers Jahrhunderts eine Entwicklung vollständig erlangt haben werde, welche eben so wichtig und noch bedeutender sein dürfte, als diejenige, welche der glänzenden Epoche des 18. Jahrhunderts zum Ruhme gereicht.

Herr General *Baeyer* fügt einige Worte hinzu, um an die unsterblichen Arbeiten zu erinnern, welche Frankreich zur Bestimmung der Gestalt und Grösse der Erde ausgeführt hat, und welche mehreren Generationen als Muster gedient haben. Herr *Baeyer* schätzt sich glücklich, den Zeitpunkt erlebt zu haben, wo die wissenschaftliche Association, welche er hervorgerufen hat, thatsächlich ganz Europa umfasst und wo man mit Zuversicht hoffen darf, dass das hohe Ziel unseres gemeinsamen Werkes erreicht werden wird.

Design Process





Different Types of Users with Different Requirements

- **Public Users:** access to a curated and accessible corpus and corpus interpretation
- **Research Scholars:** possibility of easily obtaining text analysis data on any part of the corpus
- **NLP / Programming-savvy Scholars:** access to raw data and low-level computable representation of the corpus
- **DH Specialist:** creation of a reusable and adaptable “stack” for other text-analysis projects

Catering to All Levels of Tech Savvy

- **Public Users:** Simplified interface allowing users to navigate the website
- **Scholars:** Collaborative work with DH Specialist to perform text analysis (importance of export formats)
- **Other DH Specialists:** Easy access to data through Jupyter Notebook and helper functions on models

General User Experience Design Focus

- Allow researchers to zero-in and isolate the parts of the corpus they are interested in
- Consider the public-facing part of the platform from the start
- Facilitate access to digitized versions of the original material (IIIF)

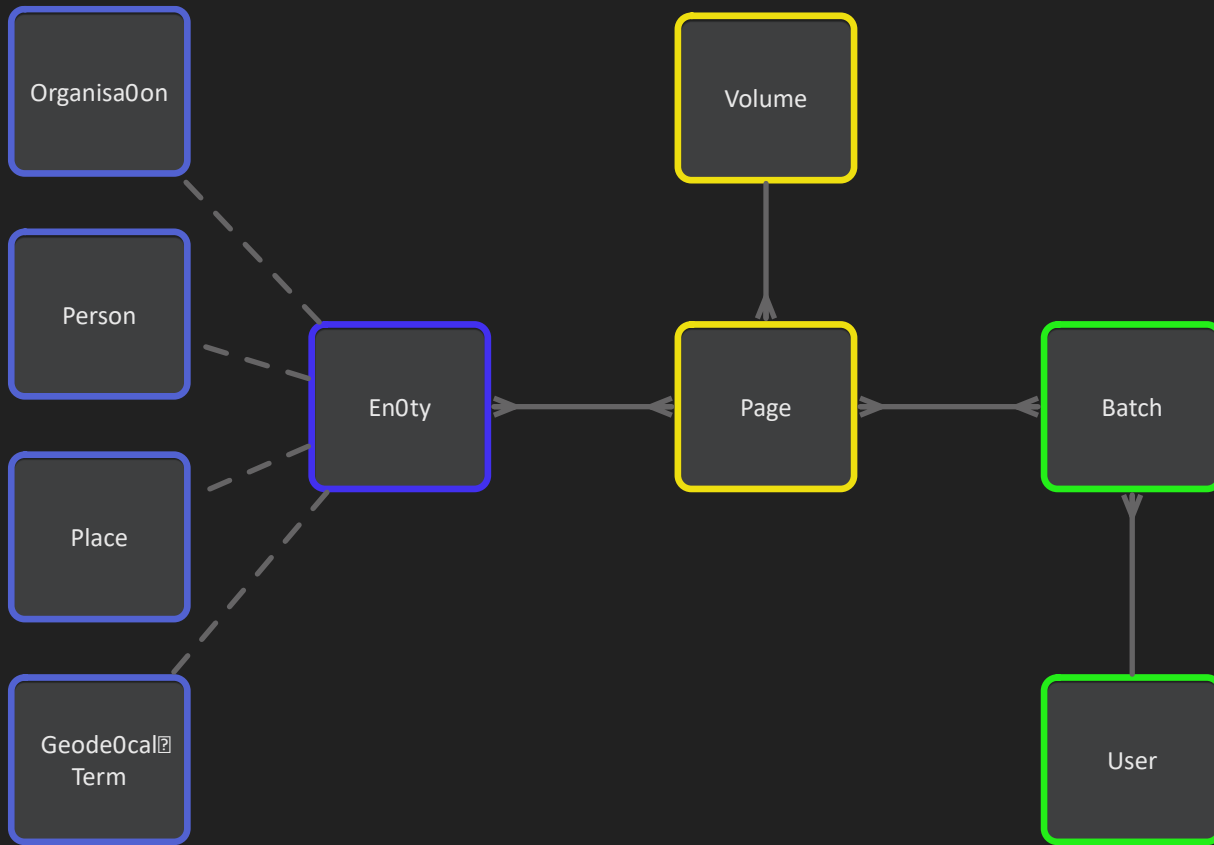
Implementation

What We Implemented

- A “public” side allowing the general public to explore the corpus
- A “private” corpus exploration browser interface for research
- A “power-user” section allowing experimentation and the use of advanced NLP libraries of the corpus using Jupyter Notebook

A Quick Note on OCR

- First Rough Run: Tesseract used on own computer
- Tried: ABBY (not very successful)
- Second (Cleaner) Run: Language detection + Computer-Vision assisted Tesseract



General Features



Search

Query:

Volume Date Range:

1868 - 1912

Any language

Any section type

Any feature

Search

Batches

Current Batch:

1868-1912 fr

Create a new batch from this selection of pages:

1862-1912

Create new batch

Search results

13335 pages found

« 1 2 3 4 5 6 7 8 »

1862 > 0001

language: French - score: 1.0

section: Generalbericht

feature: None

View image

1862 > 0002

language: English - score: 0.93

section: Generalbericht

feature: None

View image

1862 > 0003

language: Italian - score: 0.86

section: Generalbericht

feature: None

View image

1862 > 0004

language: English - score: 0.99

section: Generalbericht

feature: None

View image

1862 > 0005

language: German - score: 1.0

section: Generalbericht

feature: None

View image

1862 > 0006

language: German - score: 1.0

section: Generalbericht

feature: None

View image

1862 > 0007

language: German - score: 1.0

section: Generalbericht

feature: None

View image

Search

Query:

Volume Date Range:

1862 - 1912

Any language

Any section type

Any feature

 Search

Batches

Current Batch:

1862-1912 Generalbericht

Query Syntax

The following operators are supported:

- **+** signifies AND operation
- **|** signifies OR operation
- **-** negates a single token
- **"** wraps a number of tokens to signify a phrase for searching
- ***** at the end of a term signifies a prefix query
- **(and)** signify precedence
- **~N** after a word signifies edit distance (fuzziness)
- **~N** after a phrase signifies slop amount (the number of other words permitted between words)

To use one of these characters literally, escape it with a preceding backslash (`\`).

Close

1868 > 0036

General-Bericht über

id: 1868 key: Z

url: http://gfzpublic.gfz

date: 1869 plac

publisher: Reimer

language: German - s

tn und ist vor eini

solide Mauerwerk

ruhenden Pyrami

Beobachter befes

Fussboden des Th

jederzeit die Stellung des Instruments sowohl, als eines im Thurme aufgestellten Heliotrops auf die Marke beziehen, so dass die Lage des

Thurmknopfs durchaus nicht in Betracht kommt. Auf solche Weise wird man den Thurm mit Sicherheit als Gradmessungs-Station benutzen

können. In Lauenburg ist das Schumacher'sche Stations- Centrum jetzt auch nicht mehr vorhanden. An der Stelle dieses Dreieckspunkts wurde,

ein Paar Fuss unter der Erdoberfläche, nur noch ein viereckter eichener Pfahl vorgefunden, dessen Kopf ganz verfault war und keine Marke

enthielt. Letztere ist wahrscheinlich beim Wegräumen von Paschen's Gerüst verloren gegangen. Herr Geheime- Rath Paschen hat nämlich im Jahre

1854 die Marke noch vorgefunden und nahe bei derselben drei mit Bohrlöchern und Bleimarken versehene Granitsteine circa 2 Fuss tief in die

Erde ver- senkt, und die Lage der ersten Marke gegen die letztern durch gemessene Distanzen und Richtungen scharf bestimmt. Nach diesen

Angaben liess ich das Schumacher'sche Stations- Centrum wieder herstellen. Granitpfeiler habe ich bereits in Lauenburg, Varendorf, Niendorf,

Rahlstedt und Hohenhorst aufstellen lassen. An jedem dieser Punkte geschah die Aufstellung in folgender Weise: Auf einem Fundament von

Ziegelsteinen wurde eine Steinplatte, deren obere Fläche nahezu zwei Fuss unter der Erdoberfläche sich befindet, mit Cement befestigt. In der

8 19 »

 View image

3%BCber%20die%20Europ%C3%A4ische%20Gradmessung%20f%C3%BCr%20da

uthor: Bruhns, Carl

language: German

num_pages: 68

m wieder aufgebaut. : Das dem Anscheine nach sehr

r werde ich an dem Gebälk der auf dem Mauerwerk

dern Theile des Gebälks einen Fussboden für den

struments befinden wird, kann man ein Loth bis zum

k mit eingegossener Marke festgelegt, so lässt sich

jederzeit die Stellung des Instruments sowohl, als eines im Thurme aufgestellten Heliotrops auf die Marke beziehen, so dass die Lage des

Thurmknopfs durchaus nicht in Betracht kommt. Auf solche Weise wird man den Thurm mit Sicherheit als Gradmessungs-Station benutzen

können. In Lauenburg ist das Schumacher'sche Stations- Centrum jetzt auch nicht mehr vorhanden. An der Stelle dieses Dreieckspunkts wurde,

ein Paar Fuss unter der Erdoberfläche, nur noch ein viereckter eichener Pfahl vorgefunden, dessen Kopf ganz verfault war und keine Marke

enthielt. Letztere ist wahrscheinlich beim Wegräumen von Paschen's Gerüst verloren gegangen. Herr Geheime- Rath Paschen hat nämlich im Jahre

1854 die Marke noch vorgefunden und nahe bei derselben drei mit Bohrlöchern und Bleimarken versehene Granitsteine circa 2 Fuss tief in die

Erde ver- senkt, und die Lage der ersten Marke gegen die letztern durch gemessene Distanzen und Richtungen scharf bestimmt. Nach diesen

Angaben liess ich das Schumacher'sche Stations- Centrum wieder herstellen. Granitpfeiler habe ich bereits in Lauenburg, Varendorf, Niendorf,

Rahlstedt und Hohenhorst aufstellen lassen. An jedem dieser Punkte geschah die Aufstellung in folgender Weise: Auf einem Fundament von

Ziegelsteinen wurde eine Steinplatte, deren obere Fläche nahezu zwei Fuss unter der Erdoberfläche sich befindet, mit Cement befestigt. In der



Quit

Logout

Files

Running

Clusters

Select items to perform actions on them.

Upload

New ▾

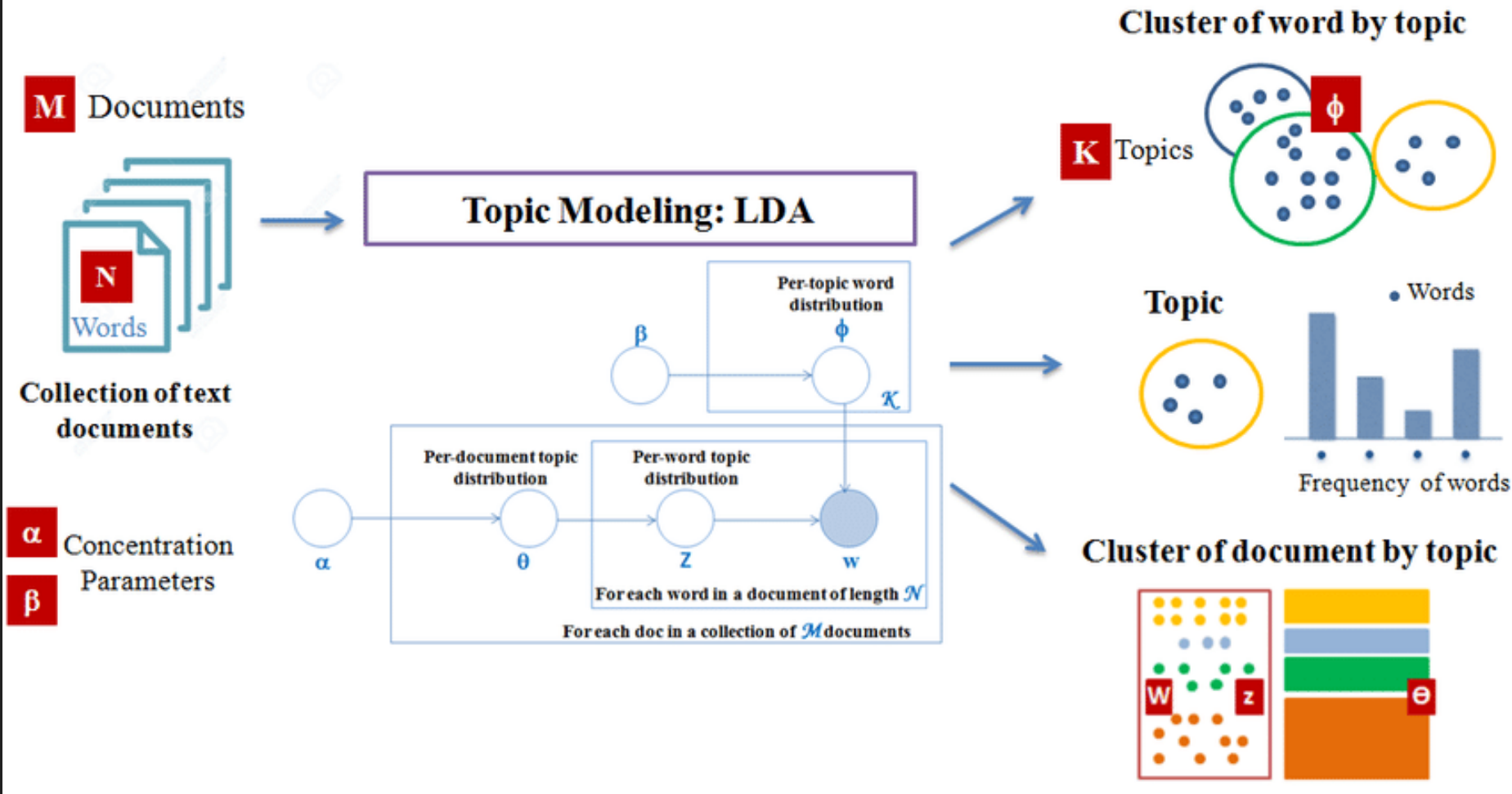


<input type="checkbox"/> 0 ▾	📁 / NLP	Name ▾	Last Modified	File size
	📁 ..		seconds ago	
<input type="checkbox"/>	📁 tf_idf_output		3 months ago	
<input type="checkbox"/>	📄 Batch Entity Co-occurrence.ipynb		9 months ago	41.8 kB
<input type="checkbox"/>	📄 Co-occurrence.ipynb		2 years ago	10.6 kB
<input type="checkbox"/>	📄 NER.ipynb		2 years ago	425 kB
<input type="checkbox"/>	📄 TF-IDF-ENGLISH.ipynb		9 months ago	1.05 MB
<input type="checkbox"/>	📄 TF-IDF-sklearn.ipynb		9 months ago	271 kB
<input type="checkbox"/>	📄 TF-IDF-sklearn_EN_DE_COMPARISON.ipynb	Running	9 days ago	75.2 kB
<input type="checkbox"/>	📄 TF-IDF-sklearn_ENGLISH_COMPARISONS.ipynb		4 months ago	496 kB
<input type="checkbox"/>	📄 TF-IDF-sklearn_GERMAN_COMPARAISONS.ipynb		3 months ago	56.2 kB
<input type="checkbox"/>	📄 TF-IDF.ipynb		9 months ago	1.95 MB
<input type="checkbox"/>	📄 Tokenization-English.ipynb		5 months ago	10 kB
<input type="checkbox"/>	📄 Tokenization.ipynb		2 years ago	12.7 kB
<input type="checkbox"/>	📄 allfiles-20211105-1109.tar.gz		3 months ago	34.7 MB
<input type="checkbox"/>	📄 count_for_9		4 months ago	279 kB
<input type="checkbox"/>	📄 edges.csv		2 years ago	50.7 kB
<input type="checkbox"/>	📄 tfidf_outputs-20211029-1152 tar.gz		4 months ago	8.19 MB

Main NLP Methods Implemented

- Topic Modelling
- TF-IDF
- Named Entity Recognition

- Multilingual Approach:
 - Automatic Translation of the whole corpus for analysis in one single language



Multilingual topic modelling for tracking COVID-19 trends based on Facebook data analysis - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/LDA-topic-modelling-process_fig1_343176425 [accessed 7 Feb, 2022]

```
[ ] pyLDAvis.sklearn.prepare(lda_tf, dtm_tf, tf_vectorizer)
```

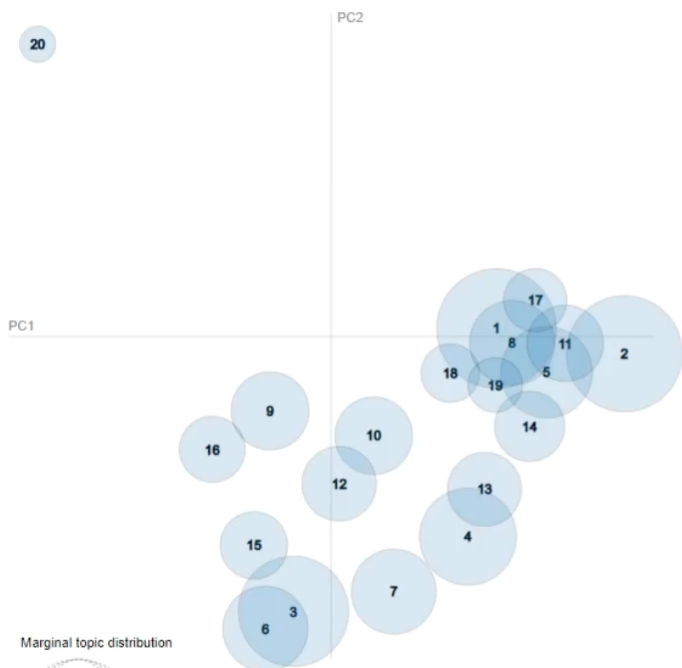
Selected Topic:

Slide to adjust relevance metric:⁽²⁾

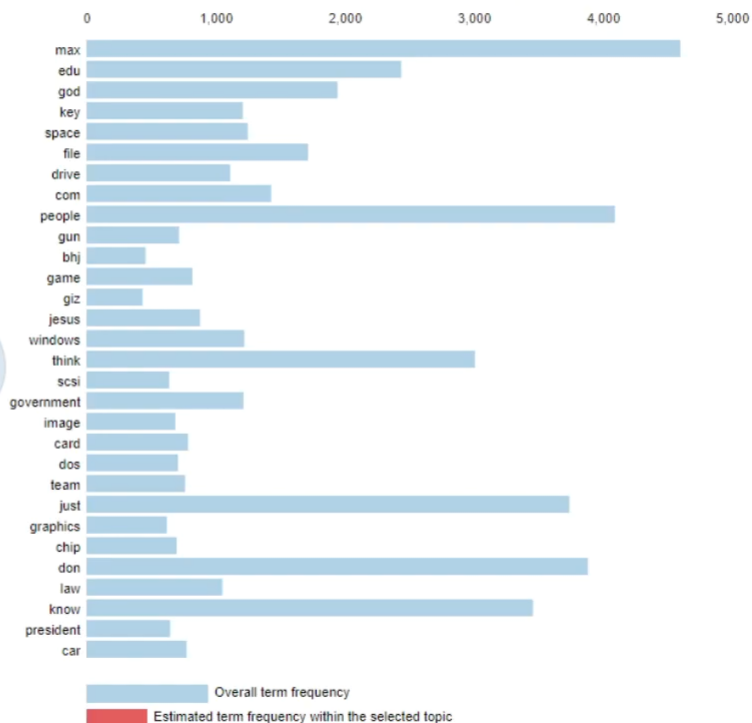
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms¹



- saliency(term w) = frequency(w) * $[\sum_t p(t | w) * \log(p(t | w)/p(t))]$ for topics t ; see Chuang et. al (2012)
- relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

```
'Bericht',  
'jetz',  
'Diese'.
```

Prepare the token lists and create the dictionaries

```
In [9]: en_doc_per_volume = []  
de_doc_per_volume = []  
doc_dictionary = []  
for volume, tokens in en_final_tokens_per_volume.items():  
    en_doc_per_volume.append(' '.join(tokens))  
    doc_dictionary.append(volume)  
  
for volume, tokens in de_final_tokens_per_volume.items():  
    de_doc_per_volume.append(' '.join(tokens))  
    doc_dictionary.append(volume)
```

Run the TFIDF proper:

max_df is used for removing terms that appear too frequently, also known as "corpus-specific stop words".

For example:

max_df = 0.50 means "ignore terms that appear in more than 50% of the documents". max_df = 25 means "ignore terms that appear in more than 25 documents". The default max_df is 1.0, which means "ignore terms that appear in more than 100% of the documents". Thus, the default setting does not ignore any terms.

min_df is used for removing terms that appear too infrequently.

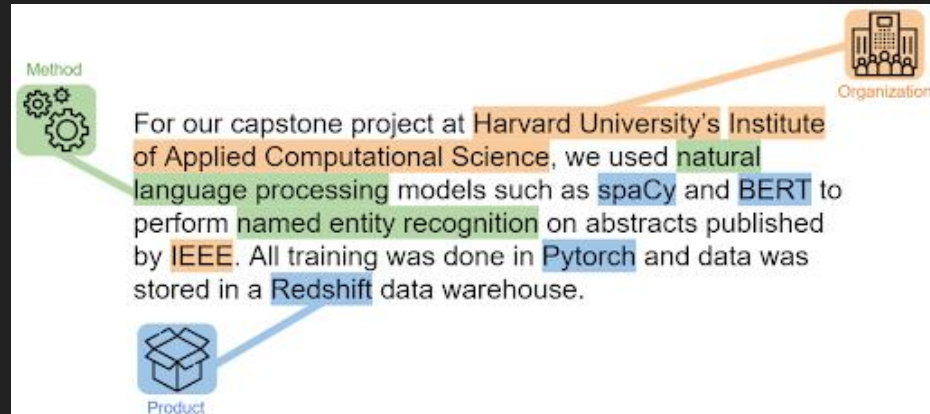
For example:

min_df = 0.01 means "ignore terms that appear in less than 1% of the documents". min_df = 5 means "ignore terms that appear in less than 5 documents". The default min_df is 1, which means "ignore terms that appear in less than 1 document". Thus, the default setting does not ignore any terms.

```
In [11]: # HYPERPARAMS APPLIED TO BOTH VECTORIZERS  
# PLEASE CHANGE MAX_DF AND MIN_DF HERE:  
max_df=.60  
min_df=.30  
  
en_vectorizer = TfidfVectorizer(max_df=max_df, min_df=min_df, stop_words=None, use_idf=True, norm=None)  
en_transformed_documents = en_vectorizer.fit_transform(en_doc_per_volume)
```

Named Entity Recognition (NER)

A process through which a sentence or a chunk of text is parsed through to find entities that can be put under categories like names, organizations, locations, etc.



NER Model Training

- Annotation with Prodigy
- Model training using Prodigy & Spacy

- First approach: one model for the whole corpus
- Second approach: one model per language

Entities

- Persons
- Organisations
- Places
- Terms

Annotate

- German
- French
- Italian
- English

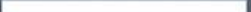
prodigy

PROJECT INFO

DATASET	fr_test
LANGUAGE	fr
RECIPE	ner.manual
VIEW ID	ner_manual

PROGRESS

THIS SESSION	1
TOTAL	1

 ∞

HISTORY

129 ("Yrayuq ayeis) ×

© 2017-2022 Explosion (Prodigy v1.11.4)

PER 1 LOC 2 ORG 3 GEO_TERM 4

beaucoup la hauteur de la colonne de mercure. Par des changements dans la construction ↵

nous pûmes réduire cette influence à un minimum. ↵ ↵

Le diamètre intérieur de la partie supérieure du tube dans les baromètres à lecture ↵

ordinaire était de 8 mm. Afin de diminuer l'influence des vibrations possibles du navire, ↵





les baromètres étaient attachés au moyen d'une suspension à Cardan à une plaque en fer ↵

qui était attachée au moyen de trois fils en acier de 1 mm. d'épaisseur à une seconde ↵

suspension à Cardan fixée à un support. Afin que les oscillations de la plaque en fer et des ↵

baromètres qui y étaient attachés ne devinssent pas trop fortes, un ressort en sp

du milieu de

    ↵ ↵

Co-occurrence Network

Co-occurrence network, sometimes referred to as a semantic network, is a method to analyse text that includes a graphic visualization of potential relationships between people, organizations, concepts, or other entities represented within written material.



Entities

Persons

Organisations

Places

Terms

Annotate

German

French

Italian

English

30600 Persons

« 10 11 12 13 14 15 16 17 »

Persons detected through NER

Text	Pages
L'Hardy	18
IV	18
Bakhuysen	18
Id	18
Maass	18
Haller	18
Wittstein	18
Lorieux	18
Tenner	18
Dr. Baeyer	18
Tacchini	18
Nahlik	18
m. F.	18
parallele	18
von Zachariae	17
v. Kalman	17

Reflections

Why do we often end up building a new platform?

- Technology evolves
- It is usually easier to build something from scratch than change an existing platform to fit the project's needs
- The design of an analysis tool is dependent on:
 - the nature and structure of the corpus
 - The research question(s)

If we always end up building a new platform, how can we produce software that facilitates this?

- Focus on "developer-centred stack"
- Extraction of useful created software as libraries?
- Use of Jupyter notebooks for analysis algorithms that allows for experimentation

Easy to Configure & Use



Works with any Corpus

Has Powerful Text Analysis Features

Take-aways

- Consider the limited lifespan of pipelines and platforms
- Facilitate access to the corpus (in any form) for any level of tech-savvy
- ML-generated entities need a lot of manual curation and further model training
- Focus on research data creation and preservation
 - Consider what data will be produced from the start of the project
 - Importance of Linked Data Standards and ontologies / conceptual reference models such as CIDOC CRM

Lessons Learnt

- Don't underestimate how much resources and help you might need for OCR and data cleaning
- One digital humanities specialist alone is not enough for such an ambitious project
- Focusing on many types of analysis at once results in a somewhat shallow approach as opposed to focussing on one or two particular approaches

Future Work

- MTE project ended about a month ago
- Still working on making this NLP platform better
- I would like to load more corpuses in it and improve its functionalities
- Once cleaned up, publish the source code on GitHub
- Please get in touch with me at pbelouin@mpiwg-berlin.mpg.de if you would like to give it a go!